



or as a tableau,

$$\begin{aligned} \left( \begin{array}{c} \mathbf{U}^T \end{array} \right) \cdot \left( \begin{array}{c} \mathbf{U} \end{array} \right) &= \left( \begin{array}{c} \mathbf{V}^T \end{array} \right) \cdot \left( \begin{array}{c} \mathbf{V} \end{array} \right) \\ &= \left( \begin{array}{c} \mathbf{1} \end{array} \right) \end{aligned} \quad (2.6.4)$$

Since  $\mathbf{V}$  is square, it is also row-orthonormal,  $\mathbf{V} \cdot \mathbf{V}^T = \mathbf{1}$ .

The SVD decomposition can also be carried out when  $M < N$ . In this case the singular values  $w_j$  for  $j = M + 1, \dots, N$  are all zero, and the corresponding columns of  $\mathbf{U}$  are also zero. Equation (2.6.2) then holds only for  $k, n \leq M$ .

The decomposition (2.6.1) can always be done, no matter how singular the matrix is, and it is “almost” unique. That is to say, it is unique up to (i) making the same permutation of the columns of  $\mathbf{U}$ , elements of  $\mathbf{W}$ , and columns of  $\mathbf{V}$  (or rows of  $\mathbf{V}^T$ ), or (ii) forming linear combinations of any columns of  $\mathbf{U}$  and  $\mathbf{V}$  whose corresponding elements of  $\mathbf{W}$  happen to be exactly equal. An important consequence of the permutation freedom is that for the case  $M < N$ , a numerical algorithm for the decomposition need not return zero  $w_j$ 's for  $j = M + 1, \dots, N$ ; the  $N - M$  zero singular values can be scattered among all positions  $j = 1, 2, \dots, N$ .

At the end of this section, we give a routine, `svdcmp`, that performs SVD on an arbitrary matrix  $\mathbf{A}$ , replacing it by  $\mathbf{U}$  (they are the same shape) and returning  $\mathbf{W}$  and  $\mathbf{V}$  separately. The routine `svdcmp` is based on a routine by Forsythe et al. [1], which is in turn based on the original routine of Golub and Reinsch, found, in various forms, in [2-4] and elsewhere. These references include extensive discussion of the algorithm used. As much as we dislike the use of black-box routines, we are going to ask you to accept this one, since it would take us too far afield to cover its necessary background material here. Suffice it to say that the algorithm is very stable, and that it is very unusual for it ever to misbehave. Most of the concepts that enter the algorithm (Householder reduction to bidiagonal form, diagonalization by  $QR$  procedure with shifts) will be discussed further in Chapter 11.

If you are as suspicious of black boxes as we are, you will want to verify yourself that `svdcmp` does what we say it does. That is very easy to do: Generate an arbitrary matrix  $\mathbf{A}$ , call the routine, and then verify by matrix multiplication that (2.6.1) and (2.6.4) are satisfied. Since these two equations are the only defining requirements for SVD, this procedure is (for the chosen  $\mathbf{A}$ ) a complete end-to-end check.

Now let us find out what SVD is good for.

## SVD of a Square Matrix

If the matrix  $\mathbf{A}$  is square,  $N \times N$  say, then  $\mathbf{U}$ ,  $\mathbf{V}$ , and  $\mathbf{W}$  are all square matrices of the same size. Their inverses are also trivial to compute:  $\mathbf{U}$  and  $\mathbf{V}$  are orthogonal, so their inverses are equal to their transposes;  $\mathbf{W}$  is diagonal, so its inverse is the diagonal matrix whose elements are the reciprocals of the elements  $w_j$ . From (2.6.1) it now follows immediately that the inverse of  $\mathbf{A}$  is

$$\mathbf{A}^{-1} = \mathbf{V} \cdot [\text{diag}(1/w_j)] \cdot \mathbf{U}^T \quad (2.6.5)$$

The only thing that can go wrong with this construction is for one of the  $w_j$ 's to be zero, or (numerically) for it to be so small that its value is dominated by roundoff error and therefore unknowable. If more than one of the  $w_j$ 's have this problem, then the matrix is even more singular. So, first of all, SVD gives you a clear diagnosis of the situation.

Formally, the *condition number* of a matrix is defined as the ratio of the largest (in magnitude) of the  $w_j$ 's to the smallest of the  $w_j$ 's. A matrix is singular if its condition number is infinite, and it is *ill-conditioned* if its condition number is too large, that is, if its reciprocal approaches the machine's floating-point precision (for example, less than  $10^{-6}$  for single precision or  $10^{-12}$  for double).

For singular matrices, the concepts of *nullspace* and *range* are important. Consider the familiar set of simultaneous equations

$$\mathbf{A} \cdot \mathbf{x} = \mathbf{b} \quad (2.6.6)$$

where  $\mathbf{A}$  is a square matrix,  $\mathbf{b}$  and  $\mathbf{x}$  are vectors. Equation (2.6.6) defines  $\mathbf{A}$  as a linear mapping from the vector space  $\mathbf{x}$  to the vector space  $\mathbf{b}$ . If  $\mathbf{A}$  is singular, then there is some subspace of  $\mathbf{x}$ , called the nullspace, that is mapped to zero,  $\mathbf{A} \cdot \mathbf{x} = 0$ . The dimension of the nullspace (the number of linearly independent vectors  $\mathbf{x}$  that can be found in it) is called the *nullity* of  $\mathbf{A}$ .

Now, there is also some subspace of  $\mathbf{b}$  that can be "reached" by  $\mathbf{A}$ , in the sense that there exists some  $\mathbf{x}$  which is mapped there. This subspace of  $\mathbf{b}$  is called the range of  $\mathbf{A}$ . The dimension of the range is called the *rank* of  $\mathbf{A}$ . If  $\mathbf{A}$  is nonsingular, then its range will be all of the vector space  $\mathbf{b}$ , so its rank is  $N$ . If  $\mathbf{A}$  is singular, then the rank will be less than  $N$ . In fact, the relevant theorem is "rank plus nullity equals  $N$ ."

What has this to do with SVD? SVD explicitly constructs orthonormal bases for the nullspace and range of a matrix. Specifically, the columns of  $\mathbf{U}$  whose same-numbered elements  $w_j$  are *nonzero* are an orthonormal set of basis vectors that span the range; the columns of  $\mathbf{V}$  whose same-numbered elements  $w_j$  are *zero* are an orthonormal basis for the nullspace.

Now let's have another look at solving the set of simultaneous linear equations (2.6.6) in the case that  $\mathbf{A}$  is singular. First, the set of *homogeneous* equations, where  $\mathbf{b} = 0$ , is solved immediately by SVD: Any column of  $\mathbf{V}$  whose corresponding  $w_j$  is zero yields a solution.

When the vector  $\mathbf{b}$  on the right-hand side is not zero, the important question is whether it lies in the range of  $\mathbf{A}$  or not. If it does, then the singular set of equations *does* have a solution  $\mathbf{x}$ ; in fact it has more than one solution, since any vector in the nullspace (any column of  $\mathbf{V}$  with a corresponding zero  $w_j$ ) can be added to  $\mathbf{x}$  in any linear combination.

If we want to single out one particular member of this solution-set of vectors as a representative, we might want to pick the one with the smallest length  $|\mathbf{x}|^2$ . Here is how to find that vector using SVD: Simply *replace*  $1/w_j$  by zero if  $w_j = 0$ . (It is not very often that one gets to set  $\infty = 0$ !) Then compute (working from right to left)

$$\mathbf{x} = \mathbf{V} \cdot [\text{diag}(1/w_j)] \cdot (\mathbf{U}^T \cdot \mathbf{b}) \quad (2.6.7)$$

This will be the solution vector of smallest length; the columns of  $\mathbf{V}$  that are in the nullspace complete the specification of the solution set.

Proof: Consider  $|\mathbf{x} + \mathbf{x}'|$ , where  $\mathbf{x}'$  lies in the nullspace. Then, if  $\mathbf{W}^{-1}$  denotes the modified inverse of  $\mathbf{W}$  with some elements zeroed,

$$\begin{aligned} |\mathbf{x} + \mathbf{x}'| &= |\mathbf{V} \cdot \mathbf{W}^{-1} \cdot \mathbf{U}^T \cdot \mathbf{b} + \mathbf{x}'| \\ &= |\mathbf{V} \cdot (\mathbf{W}^{-1} \cdot \mathbf{U}^T \cdot \mathbf{b} + \mathbf{V}^T \cdot \mathbf{x}')| \\ &= |\mathbf{W}^{-1} \cdot \mathbf{U}^T \cdot \mathbf{b} + \mathbf{V}^T \cdot \mathbf{x}'| \end{aligned} \quad (2.6.8)$$

Here the first equality follows from (2.6.7), the second and third from the orthonormality of  $\mathbf{V}$ . If you now examine the two terms that make up the sum on the right-hand side, you will see that the first one has nonzero  $j$  components only where  $w_j \neq 0$ , while the second one, since  $\mathbf{x}'$  is in the nullspace, has nonzero  $j$  components only where  $w_j = 0$ . Therefore the minimum length obtains for  $\mathbf{x}' = 0$ , q.e.d.

If  $\mathbf{b}$  is not in the range of the singular matrix  $\mathbf{A}$ , then the set of equations (2.6.6) has no solution. But here is some good news: If  $\mathbf{b}$  is not in the range of  $\mathbf{A}$ , then equation (2.6.7) can still be used to construct a “solution” vector  $\mathbf{x}$ . This vector  $\mathbf{x}$  will not exactly solve  $\mathbf{A} \cdot \mathbf{x} = \mathbf{b}$ . But, among all possible vectors  $\mathbf{x}$ , it will do the closest possible job in the least squares sense. In other words (2.6.7) finds

$$\mathbf{x} \quad \text{which minimizes} \quad r \equiv |\mathbf{A} \cdot \mathbf{x} - \mathbf{b}| \quad (2.6.9)$$

The number  $r$  is called the *residual* of the solution.

The proof is similar to (2.6.8): Suppose we modify  $\mathbf{x}$  by adding some arbitrary  $\mathbf{x}'$ . Then  $\mathbf{A} \cdot \mathbf{x} - \mathbf{b}$  is modified by adding some  $\mathbf{b}' \equiv \mathbf{A} \cdot \mathbf{x}'$ . Obviously  $\mathbf{b}'$  is in the range of  $\mathbf{A}$ . We then have

$$\begin{aligned} |\mathbf{A} \cdot \mathbf{x} - \mathbf{b} + \mathbf{b}'| &= |(\mathbf{U} \cdot \mathbf{W} \cdot \mathbf{V}^T) \cdot (\mathbf{V} \cdot \mathbf{W}^{-1} \cdot \mathbf{U}^T \cdot \mathbf{b}) - \mathbf{b} + \mathbf{b}'| \\ &= |(\mathbf{U} \cdot \mathbf{W} \cdot \mathbf{W}^{-1} \cdot \mathbf{U}^T - 1) \cdot \mathbf{b} + \mathbf{b}'| \\ &= |\mathbf{U} \cdot [(\mathbf{W} \cdot \mathbf{W}^{-1} - 1) \cdot \mathbf{U}^T \cdot \mathbf{b} + \mathbf{U}^T \cdot \mathbf{b}']| \\ &= |(\mathbf{W} \cdot \mathbf{W}^{-1} - 1) \cdot \mathbf{U}^T \cdot \mathbf{b} + \mathbf{U}^T \cdot \mathbf{b}'| \end{aligned} \quad (2.6.10)$$

Now,  $(\mathbf{W} \cdot \mathbf{W}^{-1} - 1)$  is a diagonal matrix which has nonzero  $j$  components only for  $w_j = 0$ , while  $\mathbf{U}^T \mathbf{b}'$  has nonzero  $j$  components only for  $w_j \neq 0$ , since  $\mathbf{b}'$  lies in the range of  $\mathbf{A}$ . Therefore the minimum obtains for  $\mathbf{b}' = 0$ , q.e.d.

Figure 2.6.1 summarizes our discussion of SVD thus far.

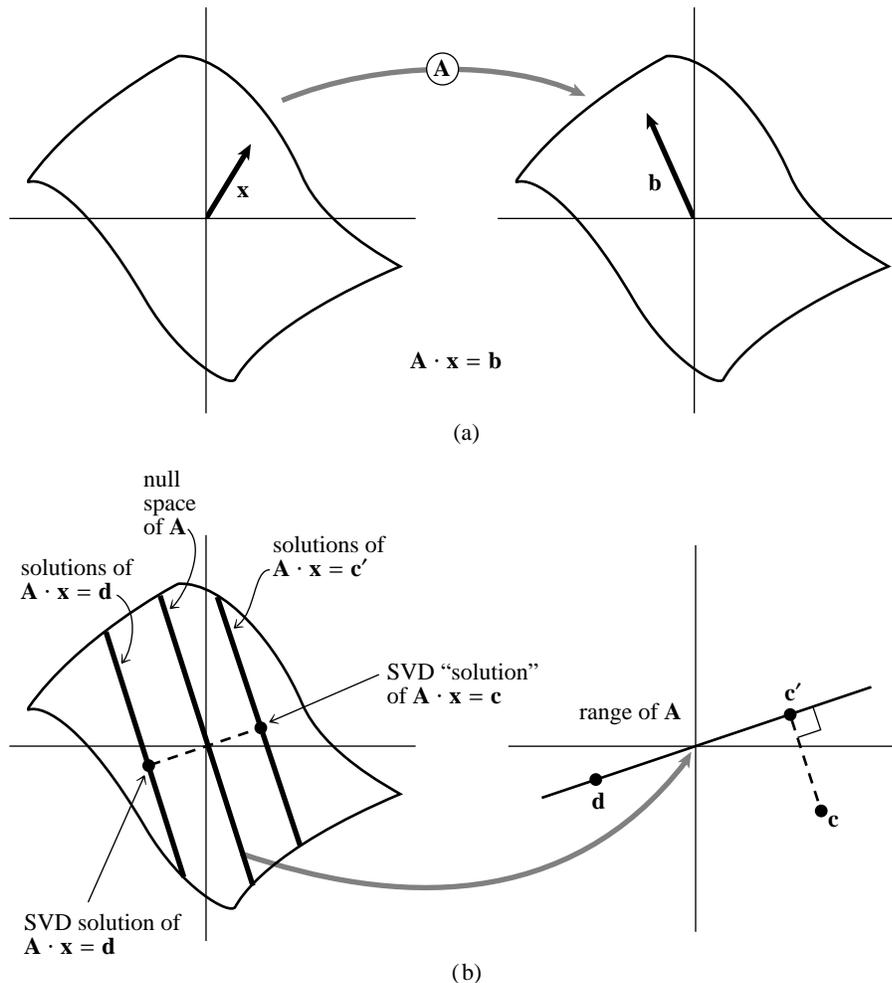


Figure 2.6.1. (a) A nonsingular matrix  $A$  maps a vector space into one of the same dimension. The vector  $x$  is mapped into  $b$ , so that  $x$  satisfies the equation  $A \cdot x = b$ . (b) A singular matrix  $A$  maps a vector space into one of lower dimensionality, here a plane into a line, called the “range” of  $A$ . The “nullspace” of  $A$  is mapped to zero. The solutions of  $A \cdot x = d$  consist of any one particular solution plus any vector in the nullspace, here forming a line parallel to the nullspace. Singular value decomposition (SVD) selects the particular solution closest to zero, as shown. The point  $c$  lies outside of the range of  $A$ , so  $A \cdot x = c$  has no solution. SVD finds the least-squares best compromise solution, namely a solution of  $A \cdot x = c'$ , as shown.

In the discussion since equation (2.6.6), we have been pretending that a matrix either is singular or else isn't. That is of course true analytically. Numerically, however, the far more common situation is that some of the  $w_j$ 's are very small but nonzero, so that the matrix is ill-conditioned. In that case, the direct solution methods of  $LU$  decomposition or Gaussian elimination may actually give a formal solution to the set of equations (that is, a zero pivot may not be encountered); but the solution vector may have wildly large components whose algebraic cancellation, when multiplying by the matrix  $A$ , may give a very poor approximation to the right-hand vector  $b$ . In such cases, the solution vector  $x$  obtained by *zeroing* the

small  $w_j$ 's and then using equation (2.6.7) is very often better (in the sense of the residual  $|\mathbf{A} \cdot \mathbf{x} - \mathbf{b}|$  being smaller) than *both* the direct-method solution *and* the SVD solution where the small  $w_j$ 's are left nonzero.

It may seem paradoxical that this can be so, since zeroing a singular value corresponds to throwing away one linear combination of the set of equations that we are trying to solve. The resolution of the paradox is that we are throwing away precisely a combination of equations that is so corrupted by roundoff error as to be at best useless; usually it is worse than useless since it “pulls” the solution vector way off towards infinity along some direction that is almost a nullspace vector. In doing this, it compounds the roundoff problem and makes the residual  $|\mathbf{A} \cdot \mathbf{x} - \mathbf{b}|$  larger.

SVD cannot be applied blindly, then. You have to exercise some discretion in deciding at what threshold to zero the small  $w_j$ 's, and/or you have to have some idea what size of computed residual  $|\mathbf{A} \cdot \mathbf{x} - \mathbf{b}|$  is acceptable.

As an example, here is a “backsubstitution” routine `svbksb` for evaluating equation (2.6.7) and obtaining a solution vector  $\mathbf{x}$  from a right-hand side  $\mathbf{b}$ , given that the SVD of a matrix  $\mathbf{A}$  has already been calculated by a call to `svdcmp`. Note that this routine presumes that *you* have already zeroed the small  $w_j$ 's. It does not do this for you. If you *haven't* zeroed the small  $w_j$ 's, then this routine is just as ill-conditioned as any direct method, and you are misusing SVD.

```

SUBROUTINE svbksb(u,w,v,m,n,mp,np,b,x)
INTEGER m,mp,n,np,NMAX
REAL b(mp),u(mp,np),v(np,np),w(np),x(np)
PARAMETER (NMAX=500)           Maximum anticipated value of n.
Solves  $A \cdot X = B$  for a vector  $X$ , where  $A$  is specified by the arrays  $u, w, v$  as returned by
svdcmp.  $m$  and  $n$  are the logical dimensions of  $a$ , and will be equal for square matrices.  $mp$ 
and  $np$  are the physical dimensions of  $a$ .  $b(1:m)$  is the input right-hand side.  $x(1:n)$  is
the output solution vector. No input quantities are destroyed, so the routine may be called
sequentially with different  $b$ 's.
INTEGER i,j,jj
REAL s,tmp(NMAX)
do 12 j=1,n                      Calculate  $U^T B$ .
  s=0.
  if(w(j).ne.0.)then              Nonzero result only if  $w_j$  is nonzero.
    do 11 i=1,m
      s=s+u(i,j)*b(i)
    enddo 11
    s=s/w(j)                       This is the divide by  $w_j$ .
  endif
  tmp(j)=s
enddo 12
do 14 j=1,n                      Matrix multiply by  $V$  to get answer.
  s=0.
  do 13 jj=1,n
    s=s+v(j,jj)*tmp(jj)
  enddo 13
  x(j)=s
enddo 14
return
END

```

Note that a typical use of `svdcmp` and `svbksb` superficially resembles the typical use of `ludcmp` and `lubksb`: In both cases, you decompose the left-hand matrix  $\mathbf{A}$  just once, and then can use the decomposition either once or many times with different right-hand sides. The crucial difference is the “editing” of the singular



given by (2.6.7), which, with nonsquare matrices, looks like this,

$$\begin{pmatrix} \mathbf{x} \end{pmatrix} = \begin{pmatrix} \mathbf{V} \end{pmatrix} \cdot \begin{pmatrix} \text{diag}(1/w_j) \end{pmatrix} \cdot \begin{pmatrix} \mathbf{U}^T \end{pmatrix} \cdot \begin{pmatrix} \mathbf{b} \end{pmatrix} \quad (2.6.12)$$

In general, the matrix  $\mathbf{W}$  will not be singular, and no  $w_j$ 's will need to be set to zero. Occasionally, however, there might be column degeneracies in  $\mathbf{A}$ . In this case you will need to zero some small  $w_j$  values after all. The corresponding column in  $\mathbf{V}$  gives the linear combination of  $\mathbf{x}$ 's that is then ill-determined even by the supposedly overdetermined set.

Sometimes, although you do not need to zero any  $w_j$ 's for *computational* reasons, you may nevertheless want to take note of any that are unusually small: Their corresponding columns in  $\mathbf{V}$  are linear combinations of  $\mathbf{x}$ 's which are insensitive to your data. In fact, you may then wish to zero these  $w_j$ 's, to reduce the number of free parameters in the fit. These matters are discussed more fully in Chapter 15.

### Constructing an Orthonormal Basis

Suppose that you have  $N$  vectors in an  $M$ -dimensional vector space, with  $N \leq M$ . Then the  $N$  vectors span some subspace of the full vector space. Often you want to construct an orthonormal set of  $N$  vectors that span the same subspace. The textbook way to do this is by Gram-Schmidt orthogonalization, starting with one vector and then expanding the subspace one dimension at a time. Numerically, however, because of the build-up of roundoff errors, naive Gram-Schmidt orthogonalization is *terrible*.

The right way to construct an orthonormal basis for a subspace is by SVD: Form an  $M \times N$  matrix  $\mathbf{A}$  whose  $N$  columns are your vectors. Run the matrix through `svdcmp`. The columns of the matrix  $\mathbf{U}$  (which in fact replaces  $\mathbf{A}$  on output from `svdcmp`) are your desired orthonormal basis vectors.

You might also want to check the output  $w_j$ 's for zero values. If any occur, then the spanned subspace was not, in fact,  $N$  dimensional; the columns of  $\mathbf{U}$  corresponding to zero  $w_j$ 's should be discarded from the orthonormal basis set.

(QR factorization, discussed in §2.10, also constructs an orthonormal basis, see [5].)

### Approximation of Matrices

Note that equation (2.6.1) can be rewritten to express any matrix  $A_{ij}$  as a sum of outer products of columns of  $\mathbf{U}$  and rows of  $\mathbf{V}^T$ , with the "weighting factors" being the singular values  $w_j$ ,

$$A_{ij} = \sum_{k=1}^N w_k U_{ik} V_{jk} \quad (2.6.13)$$

Sample page from NUMERICAL RECIPES IN FORTRAN 77: THE ART OF SCIENTIFIC COMPUTING (ISBN 0-521-43064-X)  
 Copyright (C) 1986-1992 by Cambridge University Press. Programs Copyright (C) 1986-1992 by Numerical Recipes Software.  
 Permission is granted for internet users to make one paper copy for their own personal use. Further reproduction, or any copying of machine-readable files (including this one), to any server computer, is strictly prohibited. To order Numerical Recipes books or CDROMs, visit website <http://www.nr.com> or call 1-800-872-7423 (North America only), or send email to [directcustserv@cambridge.org](mailto:directcustserv@cambridge.org) (outside North America).

If you ever encounter a situation where *most* of the singular values  $w_j$  of a matrix  $\mathbf{A}$  are very small, then  $\mathbf{A}$  will be well-approximated by only a few terms in the sum (2.6.13). This means that you have to store only a few columns of  $\mathbf{U}$  and  $\mathbf{V}$  (the same  $k$  ones) and you will be able to recover, with good accuracy, the whole matrix.

Note also that it is very efficient to multiply such an approximated matrix by a vector  $\mathbf{x}$ : You just dot  $\mathbf{x}$  with each of the stored columns of  $\mathbf{V}$ , multiply the resulting scalar by the corresponding  $w_k$ , and accumulate that multiple of the corresponding column of  $\mathbf{U}$ . If your matrix is approximated by a small number  $K$  of singular values, then this computation of  $\mathbf{A} \cdot \mathbf{x}$  takes only about  $K(M + N)$  multiplications, instead of  $MN$  for the full matrix.

## SVD Algorithm

Here is the algorithm for constructing the singular value decomposition of any matrix. See §11.2–§11.3, and also [4-5], for discussion relating to the underlying method.

```

SUBROUTINE svdcmp(a,m,n,mp,np,w,v)
  INTEGER m,mp,n,np,NMAX
  REAL a(mp,np),v(np,np),w(np)
  PARAMETER (NMAX=500)           Maximum anticipated value of n.
C  USES pythag
  Given a matrix a(1:m,1:n), with physical dimensions mp by np, this routine computes its
  singular value decomposition,  $A = U \cdot W \cdot V^T$ . The matrix  $U$  replaces a on output. The
  diagonal matrix of singular values  $W$  is output as a vector w(1:n). The matrix  $V$  (not the
  transpose  $V^T$ ) is output as v(1:n,1:n).
  INTEGER i,its,j,jj,k,l,nm
  REAL anorm,c,f,g,h,s,scale,x,y,z,rv1(NMAX),pythag
  g=0.0                           Householder reduction to bidiagonal form.
  scale=0.0
  anorm=0.0
  do 25 i=1,n
    l=i+1
    rv1(i)=scale*g
    g=0.0
    s=0.0
    scale=0.0
    if(i.le.m)then
      do 11 k=i,m
        scale=scale+abs(a(k,i))
      enddo 11
      if(scale.ne.0.0)then
        do 12 k=i,m
          a(k,i)=a(k,i)/scale
          s=s+a(k,i)*a(k,i)
        enddo 12
        f=a(i,i)
        g=-sign(sqrt(s),f)
        h=f*g-s
        a(i,i)=f-g
        do 15 j=l,n
          s=0.0
          do 13 k=i,m
            s=s+a(k,i)*a(k,j)
          enddo 13
          f=s/h
          do 14 k=i,m
            a(k,j)=a(k,j)+f*a(k,i)
          enddo 14

```

```

        enddo 15
        do 16 k=i,m
            a(k,i)=scale*a(k,i)
        enddo 16
    endif
endif
w(i)=scale *g
g=0.0
s=0.0
scale=0.0
if((i.le.m).and.(i.ne.n))then
    do 17 k=1,n
        scale=scale+abs(a(i,k))
    enddo 17
    if(scale.ne.0.0)then
        do 18 k=1,n
            a(i,k)=a(i,k)/scale
            s=s+a(i,k)*a(i,k)
        enddo 18
        f=a(i,1)
        g=-sign(sqrt(s),f)
        h=f*g-s
        a(i,1)=f-g
        do 19 k=1,n
            rv1(k)=a(i,k)/h
        enddo 19
        do 23 j=1,m
            s=0.0
            do 21 k=1,n
                s=s+a(j,k)*a(i,k)
            enddo 21
            do 22 k=1,n
                a(j,k)=a(j,k)+s*rv1(k)
            enddo 22
        enddo 23
        do 24 k=1,n
            a(i,k)=scale*a(i,k)
        enddo 24
    endif
endif
anorm=max(anorm,(abs(w(i))+abs(rv1(i))))
enddo 25
do 32 i=n,1,-1
    Accumulation of right-hand transformations.
    if(i.lt.n)then
        if(g.ne.0.0)then
            do 26 j=1,n
                Double division to avoid possible underflow.
                v(j,i)=(a(i,j)/a(i,1))/g
            enddo 26
            do 29 j=1,n
                s=0.0
                do 27 k=1,n
                    s=s+a(i,k)*v(k,j)
                enddo 27
                do 28 k=1,n
                    v(k,j)=v(k,j)+s*v(k,i)
                enddo 28
            enddo 29
        endif
        do 31 j=1,n
            v(i,j)=0.0
            v(j,i)=0.0
        enddo 31
    endif
    v(i,i)=1.0

```

Sample page from NUMERICAL RECIPES IN FORTRAN 77: THE ART OF SCIENTIFIC COMPUTING (ISBN 0-521-43064-X)  
 Copyright (C) 1986-1992 by Cambridge University Press. Programs Copyright (C) 1986-1992 by Numerical Recipes Software.  
 Permission is granted for internet users to make one paper copy for their own personal use. Further reproduction, or any copying of machine-readable files (including this one), to any server computer, is strictly prohibited. To order Numerical Recipes books or CDROMs, visit website <http://www.nr.com> or call 1-800-872-7423 (North America only), or send email to [directcustserv@cambridge.org](mailto:directcustserv@cambridge.org) (outside North America).

```

    g=rv1(i)
    l=i
  enddo 32
do 39 i=min(m,n),1,-1          Accumulation of left-hand transformations.
  l=i+1
  g=w(i)
  do 33 j=1,n
    a(i,j)=0.0
  enddo 33
  if(g.ne.0.0)then
    g=1.0/g
    do 36 j=1,n
      s=0.0
      do 34 k=1,m
        s=s+a(k,i)*a(k,j)
      enddo 34
      f=(s/a(i,i))*g
      do 35 k=i,m
        a(k,j)=a(k,j)+f*a(k,i)
      enddo 35
    enddo 36
    do 37 j=i,m
      a(j,i)=a(j,i)*g
    enddo 37
  else
    do 38 j= i,m
      a(j,i)=0.0
    enddo 38
  endif
  a(i,i)=a(i,i)+1.0
enddo 39
do 49 k=n,1,-1                Diagonalization of the bidiagonal form: Loop over
  do 48 its=1,30                singular values, and over allowed iterations.
    do 41 l=k,1,-1              Test for splitting.
      nm=l-1                    Note that rv1(1) is always zero.
      if((abs(rv1(l))+anorm).eq.anorm) goto 2
      if((abs(w(nm))+anorm).eq.anorm) goto 1
    enddo 41
    c=0.0                        Cancellation of rv1(1), if 1 > 1.
    s=1.0
    do 43 i=1,k
      f=s*rv1(i)
      rv1(i)=c*rv1(i)
      if((abs(f)+anorm).eq.anorm) goto 2
      g=w(i)
      h=pythag(f,g)
      w(i)=h
      h=1.0/h
      c= (g*h)
      s=-(f*h)
      do 42 j=1,m
        y=a(j,nm)
        z=a(j,i)
        a(j,nm)=(y*c)+(z*s)
        a(j,i)=- (y*s)+(z*c)
      enddo 42
    enddo 43
    z=w(k)
    if(l.eq.k)then              Convergence.
      if(z.lt.0.0)then          Singular value is made nonnegative.
        w(k)=-z
        do 44 j=1,n
          v(j,k)=-v(j,k)
        enddo 44
      endif
    endif
  enddo 48
enddo 49

```

Sample page from NUMERICAL RECIPES IN FORTRAN 77: THE ART OF SCIENTIFIC COMPUTING (ISBN 0-521-43064-X)  
 Copyright (C) 1986-1992 by Cambridge University Press. Programs Copyright (C) 1986-1992 by Numerical Recipes Software.  
 Permission is granted for internet users to make one paper copy for their own personal use. Further reproduction, or any copying of machine-readable files (including this one), to any server computer, is strictly prohibited. To order Numerical Recipes books or CDROMs, visit website <http://www.nr.com> or call 1-800-872-7423 (North America only), or send email to [directcustserv@cambridge.org](mailto:directcustserv@cambridge.org) (outside North America).

```

        endif
        goto 3
    endif
    if(its.eq.30) pause 'no convergence in svdcmp'
    x=w(1)          Shift from bottom 2-by-2 minor.
    nm=k-1
    y=w(nm)
    g=rv1(nm)
    h=rv1(k)
    f=((y-z)*(y+z)+(g-h)*(g+h))/(2.0*h*y)
    g=pythag(f,1.0)
    f=((x-z)*(x+z)+h*((y/(f+sign(g,f)))-h))/x
    c=1.0          Next QR transformation:
    s=1.0
    do 47 j=1,nm
        i=j+1
        g=rv1(i)
        y=w(i)
        h=s*g
        g=c*g
        z=pythag(f,h)
        rv1(j)=z
        c=f/z
        s=h/z
        f=(x*c)+(g*s)
        g=-(x*s)+(g*c)
        h=y*s
        y=y*c
        do 45 jj=1,n
            x=v(jj,j)
            z=v(jj,i)
            v(jj,j)=(x*c)+(z*s)
            v(jj,i)=-(x*s)+(z*c)
        enddo 45
        z=pythag(f,h)
        w(j)=z          Rotation can be arbitrary if z = 0.
        if(z.ne.0.0)then
            z=1.0/z
            c=f*z
            s=h*z
        endif
        f=(c*g)+(s*y)
        x=-(s*g)+(c*y)
        do 46 jj=1,m
            y=a(jj,j)
            z=a(jj,i)
            a(jj,j)=(y*c)+(z*s)
            a(jj,i)=-(y*s)+(z*c)
        enddo 46
    enddo 47
    rv1(1)=0.0
    rv1(k)=f
    w(k)=x
enddo 48
3 continue
enddo 49
return
END

```

```

FUNCTION pythag(a,b)
REAL a,b,pythag
    Computes  $(a^2 + b^2)^{1/2}$  without destructive underflow or overflow.

```

Sample page from NUMERICAL RECIPES IN FORTRAN 77: THE ART OF SCIENTIFIC COMPUTING (ISBN 0-521-43064-X)  
 Copyright (C) 1986-1992 by Cambridge University Press. Programs Copyright (C) 1986-1992 by Numerical Recipes Software.  
 Permission is granted for internet users to make one paper copy for their own personal use. Further reproduction, or any copying of machine-readable files (including this one), to any server computer, is strictly prohibited. To order Numerical Recipes books or CDROMs, visit website <http://www.nr.com> or call 1-800-872-7423 (North America only), or send email to [directcustserv@cambridge.org](mailto:directcustserv@cambridge.org) (outside North America).

```

REAL absa,absb
absa=abs(a)
absb=abs(b)
if(absa.gt.absb)then
  pythag=absa*sqrt(1.+(absb/absa)**2)
else
  if(absb.eq.0.)then
    pythag=0.
  else
    pythag=absb*sqrt(1.+(absa/absb)**2)
  endif
endif
return
END

```

(Double precision versions of `svdcmp`, `svbksb`, and `pythag`, named `dsvdcmp`, `dsvbksb`, and `dpythag`, are used by the routine `ratlsq` in §5.13. You can easily make the conversions, or else get the converted routines from the *Numerical Recipes* diskette.)

#### CITED REFERENCES AND FURTHER READING:

- Golub, G.H., and Van Loan, C.F. 1989, *Matrix Computations*, 2nd ed. (Baltimore: Johns Hopkins University Press), §8.3 and Chapter 12.
- Lawson, C.L., and Hanson, R. 1974, *Solving Least Squares Problems* (Englewood Cliffs, NJ: Prentice-Hall), Chapter 18.
- Forsythe, G.E., Malcolm, M.A., and Moler, C.B. 1977, *Computer Methods for Mathematical Computations* (Englewood Cliffs, NJ: Prentice-Hall), Chapter 9. [1]
- Wilkinson, J.H., and Reinsch, C. 1971, *Linear Algebra*, vol. II of *Handbook for Automatic Computation* (New York: Springer-Verlag), Chapter I.10 by G.H. Golub and C. Reinsch. [2]
- Dongarra, J.J., et al. 1979, *LINPACK User's Guide* (Philadelphia: S.I.A.M.), Chapter 11. [3]
- Smith, B.T., et al. 1976, *Matrix Eigensystem Routines — EISPACK Guide*, 2nd ed., vol. 6 of *Lecture Notes in Computer Science* (New York: Springer-Verlag).
- Stoer, J., and Bulirsch, R. 1980, *Introduction to Numerical Analysis* (New York: Springer-Verlag), §6.7. [4]
- Golub, G.H., and Van Loan, C.F. 1989, *Matrix Computations*, 2nd ed. (Baltimore: Johns Hopkins University Press), §5.2.6. [5]

## 2.7 Sparse Linear Systems

A system of linear equations is called *sparse* if only a relatively small number of its matrix elements  $a_{ij}$  are nonzero. It is wasteful to use general methods of linear algebra on such problems, because most of the  $O(N^3)$  arithmetic operations devoted to solving the set of equations or inverting the matrix involve zero operands. Furthermore, you might wish to work problems so large as to tax your available memory space, and it is wasteful to reserve storage for unfruitful zero elements. Note that there are two distinct (and not always compatible) goals for any sparse matrix method: saving time and/or saving space.

We have already considered one archetypal sparse form in §2.4, the band diagonal matrix. In the tridiagonal case, e.g., we saw that it was possible to save